

The American Journal of Human Genetics, Volume 92

Supplemental Data

Rare, Low-Frequency, and Common Variants in the Protein-Coding Sequence of Biological Candidate Genes from GWASs Contribute to Risk of Rheumatoid Arthritis

Dorothée Diogo, Fina Kurreeman, Eli A. Stahl, Katherine P. Liao, Namrata Gupta, Jeffrey D. Greenberg, Manuel A. Rivas, Brendan Hickey, Jason Flannick, Brian Thomson, Candace Guiducci, Stephan Ripke, Ivan Adzhubey, Anne Barton, Joel M. Kremer, Lars Alfredsson, Consortium of Rheumatology Researchers of North America, Rheumatoid Arthritis Consortium International , Shamil Sunyaev, Javier Martin, Alexandra Zhernakova, John Bowes, Steve Eyre, Katherine A. Siminovitch, Peter K. Gregersen, Jane Worthington, Lars Klareskog, Leonid Padyukov, Soumya Raychaudhuri, and Robert M. Plenge

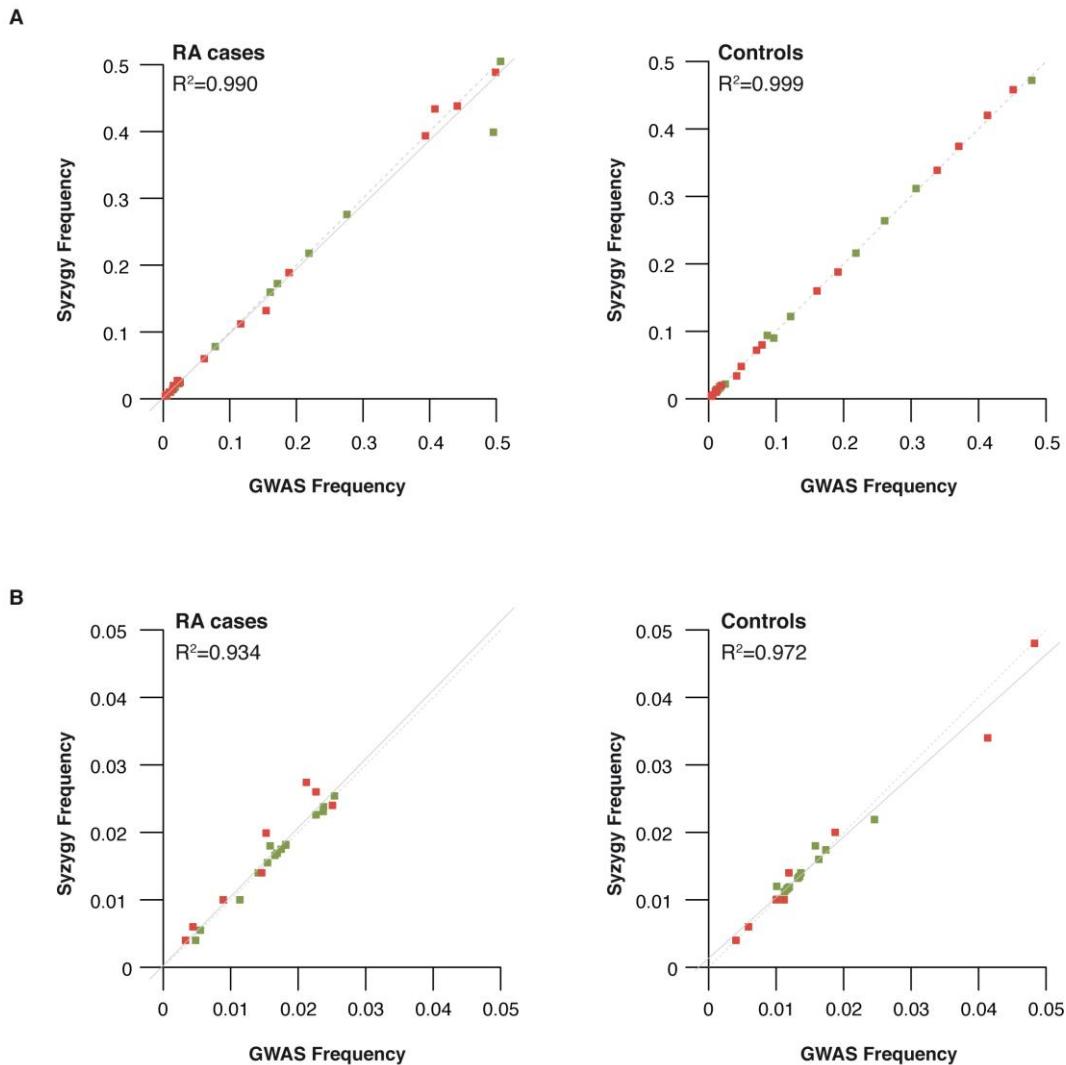


Figure S1. Correlation of Allele Frequency from Pooled Sequencing with GWAS Data

We used GWAS data available from 250 RA cases and 250 controls included in our sequencing study. The allele frequencies estimated from read counts by Syzygy and expected frequencies in the GWAS pools are shown. A. Frequencies estimates for 40 SNPs. B. Frequencies estimates for a subset of 25 SNPs with $MAF < 0.05$. SNPs genotyped and imputed in the GWAS data are shown in green and red, respectively. Pearson correlation coefficient is indicated.

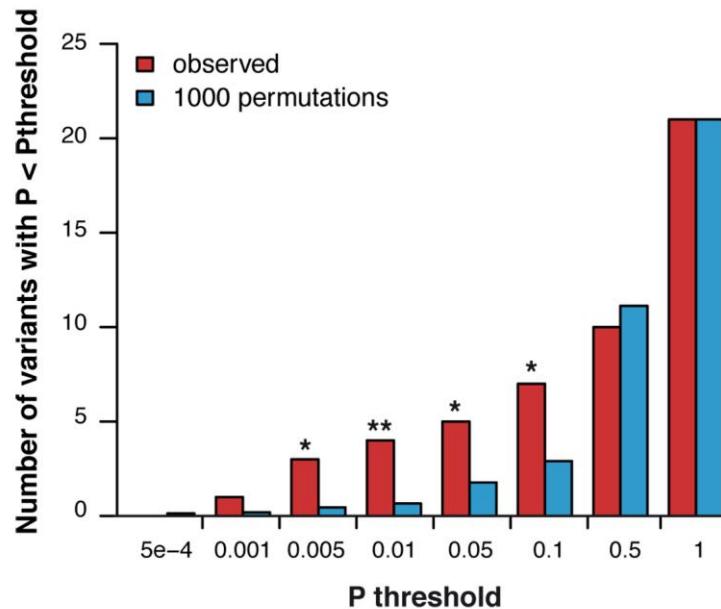


Figure S2. Enrichment of Nominal Signal of Association Driven by Nonsynonymous Variants in the Conditional Analysis, after Excluding the Known Variant in *TNFAIP3*

The number of nonsynonymous variants reaching $P < P$ threshold in our conditional analysis or after 1000 permutations of the phenotypes are shown. Significant enrichment of SNPs with $P < P$ threshold in our conditional analysis was assessed by Fisher-exact test (*, $P < 0.05$; **, $P < 0.01$).

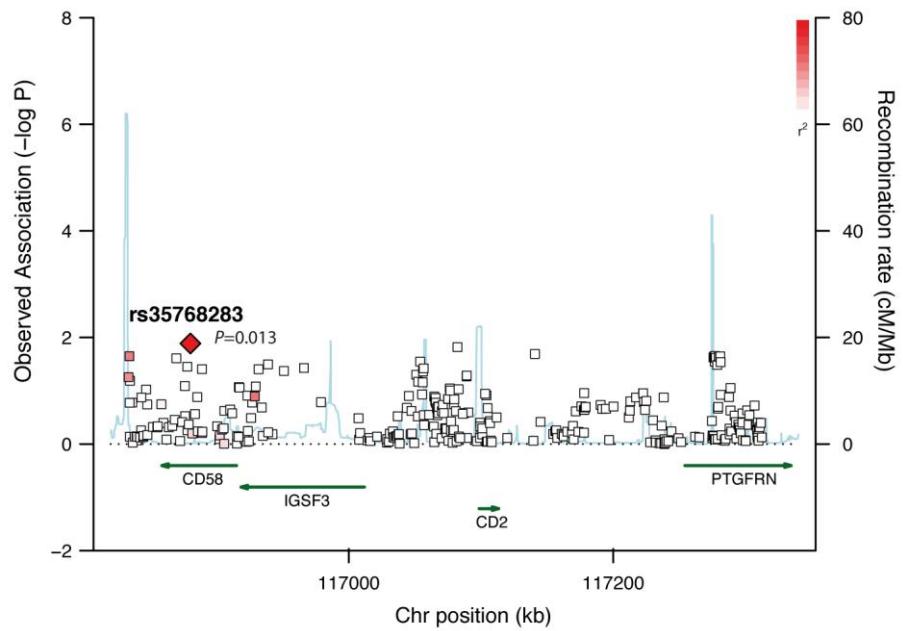


Figure S3. Absence of Evidence for a Third Signal of Association at the *CD2* Locus

Association results from conditional analysis adjusting for the best signal of association in the meta-analysis (rs624988) and the nonsynonymous variant proxy rs798036 are shown. No signal reached $P<0.01$. The best signal of association in this analysis is rs35768283, a synonymous variant in *CD58*.

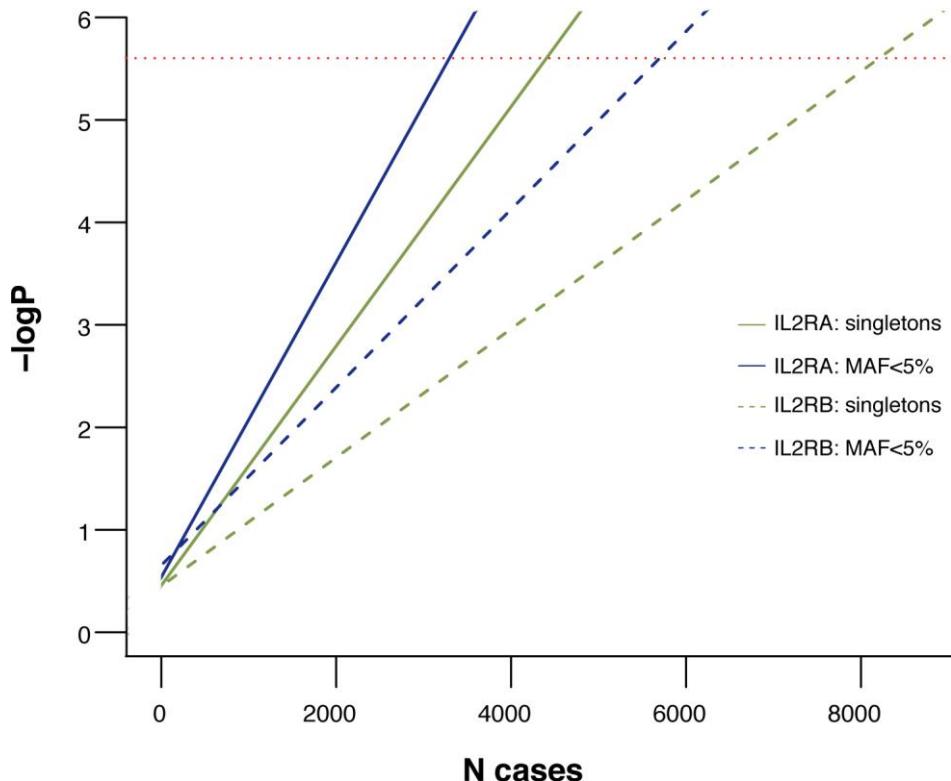


Figure S4. Estimation of Sample Size Required to Observe Significant Association to RA in *IL2RA* and *IL2RB* Gene Burden Tests

Data extrapolation from OR estimated based on counts of singletons in our RA cases and singletons described in European-American samples (the Exome Sequencing Project, <http://evs.gs.washington.edu/EVS/>) are shown in green, and data extrapolation from OR estimated based on alleles counts for variants with MAF<5% in our RA cases and our controls is shown in blue. A case:control ratio of 1 is modeled. P values were calculated using one-sided Fisher's exact test.

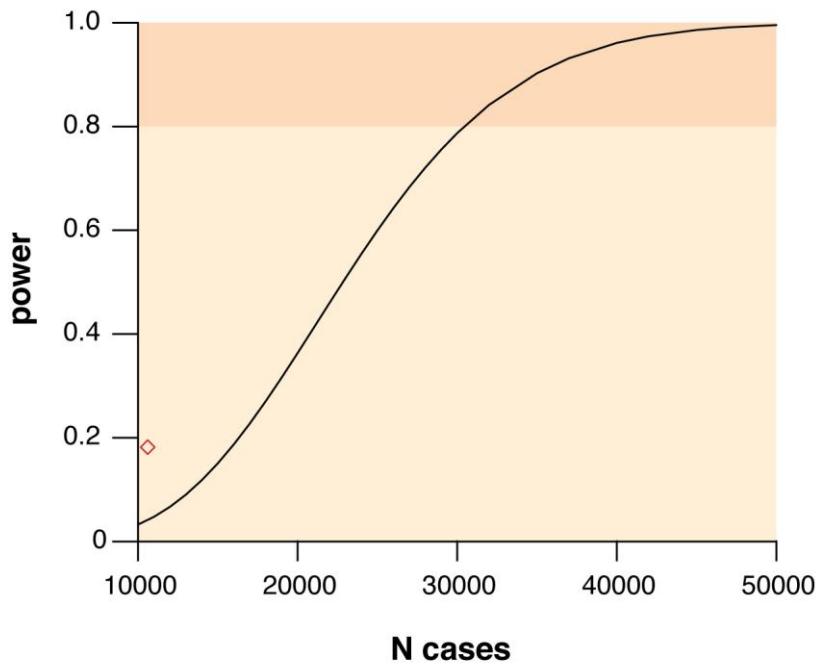


Figure S5. Estimation of Sample Size Required to Achieve Genome-wide Significance ($P=5\times 10^{-8}$) in Single SNP Analysis for the Independent SNP Found in *CD2*.

Here, we assumed a disease prevalence=0.75%, OR=1.14, risk allele frequency=0.914, and a ratio N controls/N cases=1. The power to detect association at $P=5\times 10^{-8}$ with the sample size in our study (10609 cases, 35613 controls) is indicated by a red diamond. Power >80% is highlighted in orange.

Table S2. Sensitivity of the Sequencing Results Based on GWAS Data

Number of Alleles per Pool	Total Expected ^a	Total Detected	% Detected
0	33	33	100
1	84	83	99
2	31	31	100
3	22	22	100
≥ 4	6	6	100
TOTAL	176	175	99

^a based on GWAS data

Table S3. Description of the Genotyping Samples

	Collection	ACPA+ Cases	Controls
ichip	UK	2406	8430
	Swedish EIRA	1762	1940
	US	1803	2134
	Dutch	330	2004
	Swedish Umea	524	963
	Spanish	397	399
	i2b2/corrona (US)	1024	1863
TOTAL		8246	17733
GWAS	BRASS (US)	479	1627
	Canada	586	1553
	NARAC2 (US)	746	6567
	WTCCC (UK)	552	8125
TOTAL		2363	17872
TOTAL		10609	35605

Table S4. Association Results for the 47 Coding Variants Tagged in Our Genotyping Data

							Meta-analysis							Conditional Analysis							
Coding Variant				Proxy			r^2 ^a	Proxy				Best Hit			r^2 ^b	Proxy		Best Hit			r^2 ^b
ID	Coordinates	Gene	Pr	ID	BP	m/M	MAF	N	OR	P	ID	P	P	ID	P	P	ID	P	P		
A. Missense Variants																					
rs2476601	1:114179091	PTPN22	pr_d	rs2476601	114179091	-	A/G	0.113	11	1.82 (1.77-1.87)	0	rs2476601	0	1	-	-	-	-	-	-	
rs2230926	6:138237759	TNFAIP3	B	rs2230926	138237759	-	C/A	0.038	11	1.38 (1.3-1.47)	6.8E-14	rs6920220	1.9E-18	0.03	1.4E-09	rs58721818	3.3E-10	0.87			
rs1801274	1:159746369	FCGR2A	B	rs1801274	159746369	-	A/G	0.494	11	1.1 (1.06-1.13)	2.4E-07	rs10494360	1.8E-07	0.15	5.0E-04	rs1801274	4.7E-04	1			
rs699738	1:117112670	CD2	B	rs798036	117110353	1	A/T	0.087	11	0.88 (0.82-0.95)	1.0E-04	rs624988	2.4E-08	0.02	0.0017	rs798036	0.0016	1			
rs2234163	1:2483112	TNFRSF14	B	imm_1_2488683	2488683	1	A/G	0.006	7	1.52 (1.28-1.76)	8.0E-04	rs11585048	7.4E-11	~0	0.0019	rs1886731	7.5E-04	~0			
rs55758736	8:11442985	BLK	B	rs55758736	11442985	-	A/G	0.012	11	0.8 (0.62-0.97)	0.01	rs922483	1.8E-06	~0	0.008	rs77072957	1.2E-04	~0			
rs41267765	6:159382412	TAGAP	B	rs41267765	159382412	-	A/G	0.029	11	1.18 (1.04-1.28)	0.0018	rs2451258	5.6E-10	~0	0.011	rs112904761	0.0017	0.76			
rs72650666	10:6106308	IL2RA	po_d	seq-rs36065822	6106308	-	A/G	0.002	5	1.41 (0.99-1.83)	0.11	rs10795791	2.9E-09	~0	0.08	rs11256360	5.0E-4	~0			
rs811925	6:106654065	PRDM1	B	rs486522	106656091	0.99	A/G	0.175	11	1.04 (0.99-1.08)	0.13	rs6911690	0.00012	~0	0.087	rs112096104	0.0019	~0			
rs2234167	1:2480088	TNFRSF14	B	rs10910093	2495573	1	A/G	0.139	11	1.09 (1.04-1.13)	0.001	rs11585048	7.4E-11	0.07	0.13	rs1886731	7.5E-04	0.15			
rs228942	22:35854565	IL2RB	B	rs228942	35854565	-	A/C	0.18	11	0.97 (0.92-1.01)	0.13	rs3218251	1.1E-06	~0	0.14	rs5756391	4.3E-94	~0			
rs2185379	6:106642946	PRDM1	B	rs2185379	106642946	-	A/G	0.031	11	1.03 (0.94-1.13)	0.51	rs6911690	0.00012	~0	0.44	rs112096104	0.0019	~0			
rs4915154	1:196938276	PTPRC	B	rs4915154	196938276	-	G/A	0.008	7	0.89 (0.67-1.11)	0.29	rs1932437	5.2E-08	0.02	0.59	rs114128254	3.7E-04	~0			
rs2236379	10:6567149	PRKCQ	B	rs2236379	6567149	-	A/G	0.251	11	0.99 (0.95-1.03)	0.49	rs10796045	1.9E-07	~0	0.6	rs6602705	2.0E-04				
rs231775	2:204440959	CTLA4	B	rs231775	204440959	-	G/A	0.39	11	1.09 (1.06-1.13)	6.3E-07	rs3087243	5.1E-14	0.51	0.61	rs1980422	3.9E-06	0.07			
rs11086998	20:44190931	CD40	B	rs11569329	44188319	1	A/G	0.007	7	1.06 (0.82-1.29)	0.65	rs6032662	1.9E-13	~0	0.65	-	0	0			

chr6_106660076	6:106660076	PRDM1	B	chr6_106660076	106660076	-	A/G	0.009	5	1.04 (0.74-1.34)	0.79	rs6911690	0.00012	~0	0.83	rs112096104	0.0019	~0
rs41269905	1:196935384	PTPRC	pr_d	rs41269905	196935384	-	C/G	0.015	11	0.89 (0.75-1.03)	0.1	rs1932437	5.2E-08	0.09	0.86	rs114128254	3.7E-04	~0
rs2230606	1:196942686	PTPRC	B	rs72738053	196949454	1	C/A	0.018	11	0.89 (0.76-1.03)	0.11	rs1932437	5.2E-08	0.09	0.88	rs114128254	3.7E-04	~0
rs17860852	6:167470814	CCR6	B	rs17860852	167470814	-	A/G	0.007	7	1.09 (0.87-1.31)	0.45	rs1571878	3.3E-13	~0	0.88	rs9459836	3.6E-04	~0
chr6_106650242	6:106650242	PRDM1	B	chr6_106650242	106650242	-	G/A	0.001	4	1.05 (0.31-1.8)	0.89	rs6911690	0.00012	~0	0.91	rs112096104	0.0019	~0
rs41314039	1:196943962	PTPRC	B	rs41314039	196943962	-	G/A	0.001	3	0.96 (0.23-1.69)	0.91	rs1932437	5.2E-08	~0	0.96	rs114128254	3.7E-04	~0
rs33996649	1:114196212	PTPN22	B	rs33996649	114196212	-	A/G	0.023	11	0.93 (0.81-1.05)	0.22	rs2476601	0	~0	0.98	rs3811021	2.0E-04	

B. Synonymous Variants

rs2239657	9:122711341	TRAF1	-	rs2239657	122711341	-	G/A	0.353	11	1.1 (1.07-1.14)	5.4E-08	rs10760126	2.0E-10	0.7	0.82	-	-	-
rs7540378	1:196970917	PTPRC	-	rs7540378	196970917	-	A/G	0.12	11	0.88 (0.83-0.93)	2.7E-06	rs1932437	5.2E-08	0.87	0.89	-	-	-
rs11810143	1:159747273	FCGR2A	-	rs6671753	159745332	1	A/G	0.132	11	1.14 (1.09-1.19)	1.8E-07	rs10494360	1.8E-07	0.99	-	-	-	-
rs2228150	10:6107975	IL2RA	-	rs2228150	6107975	-	A/G	0.03	11	1.25 (1.15-1.34)	6.6E-06	rs10795791	2.9E-09	0.03	6.0E-04	rs11256360	5.0E-4	0.89
rs228953	22:35861382	IL2RB	-	rs228953	35861382	-	A/G	0.431	11	0.95 (0.91-0.98)	0.0023	rs3218251	1.1E-06	~0	0.004	rs5756391	4.3E-04	~0
rs35768283	1:116880252	CD58	-	rs35768283	116880252	-	G/A	0.019	10	1.2 (1.08-1.33)	0.0043	rs624988	2.4E-08	~0	0.012	rs798036	0.0017	~0
rs35105682	6:159383209	TAGAP	-	rs35105682	159383209	-	A/G	0.028	11	1.17 (1.07-1.27)	0.0019	rs2451258	5.6E-10	0.01	0.015	rs112904761	0.0017	0.92
rs3747841	9:122715622	TRAF1	-	rs12235400	122770510	1	A/G	0.013	10	0.89 (0.72-1.06)	0.17	rs10760126	2.0E-10	0.01	0.03	rs76418192	0.0037	~0
rs17555072	10:6593077	PRKCQ	-	rs17555072	6593077	-	G/A	0.077	11	1.05 (0.98-1.11)	0.16	rs10796045	1.9E-07	0.01	0.05	rs6602705	2.0E-04	
rs2071171	6:167470032	CCR6	-	rs2071171	167470032	-	G/A	0.377	10	1.05 (1.02-1.09)	0.0049	rs1571878	3.3E-13	0.02	0.07	rs9459836	3.6E-04	~0
chr6_138243951	6:138243951	TNFAIP3	-	gw_060039	138243951	-	A/G	0.003	7	0.73 (0.27-1.2)	0.19	rs6920220	1.9E-18	~0	0.22	rs58721818	3.3E-10	~0
rs11258943	10:6538729	PRKCQ	-	rs11258943	6538729	-	A/G	0.217	11	1.03 (0.99-1.07)	0.23	rs10796045	1.9E-07	~0	0.23	rs6602705	2.0E-04	
rs17612648	1:196932540	PTPRC	-	rs17612648	196932540	-	C/G	0.014	7	0.92 (0.74-1.09)	0.32	rs1932437	5.2E-08	~0	0.24	rs114128254	3.7E-04	~0
rs41272649	2:204299653	CD28	-	rs41272649	204299653	-	A/G	0.012	11	0.99 (0.83-1.15)	0.9	rs3087243	5.1E-14	~0	0.39	rs1980422	4.0E-06	~0

rs2069763	4:123596932	IL2	-	rs2069763	123596932	-	A/C	0.34	11	0.98 (0.94-1.01)	0.19	rs78560100	2.3E-05	0.04	0.64	rs72687039	0.0014	0.05
rs3816668	8:11444002	BLK	-	rs3816668	11444002	-	A/G	0.486	11	1.03 (0.99-1.06)	0.15	rs922483	1.8E-06	0.04	0.7	rs77072957	1.2E-04	0.13
rs2230117	7:128375281	IRF5	-	rs2230117	128375281	-	G/T	0.003	7	0.98 (0.64-1.32)	0.92	rs3778754	3.2E-11	~0	0.79	rs4728150	0.0017	~0
rs4833837	4:123756413	IL21	-	rs4833837	123756413	-	G/A	0.329	11	0.98 (0.95-1.02)	0.4	rs78560100	2.3E-05	0.03	0.9	rs72687039	0.0014	0.04
rs17301189	10:6560987	PRKCQ	-	rs17301189	6560987	-	A/G	0.011	11	1.02 (0.86-1.19)	0.8	rs10796045	1.9E-07	~0	0.93	rs6602705	0.0002	
rs41314043	1:196944009	PTPRC	-	rs41314043	196944009	-	A/G	0.015	11	0.89 (0.75-1.04)	0.12	rs1932437	5.2E-08	0.09	0.96	rs114128254	3.7E-04	~0
rs2228149	10:6103514	IL2RA	-	rs12722608	6092847	1	A/C	0.035	11	1.05 (0.96-1.15)	0.26	rs10795791	2.9E-09	0.03	0.96	rs11256360	5.0E-4	~0
rs35751269	10:6560969	PRKCQ	-	rs35751269	6560969	-	A/G	0.021	11	1.01 (0.89-1.14)	0.85	rs10796045	1.9E-07	~0	0.97	rs6602705	2.0E-04	
rs1010273	6:106661718	PRDM1	-	rs1010273	106661718	-	A/G	0.1	11	0.96 (0.9-1.01)	0.13	rs6911690	0.00012	0.22	0.98	rs112096104	0.0019	~0

^a Linkage disequilibrium (LD) between the coding variant and its best proxy in the 1000Genomes, CEU population (release June 2011)

^b LD between the coding variant best proxy and the best signal of association in our Immunochip genotyping data

Table S5. Association Results at the Known GWAS Best Hits in the Meta-analysis and after Conditioning on the Coding Variants

Locus	Best Hit	Coding Variant	r^2	Best Hit N ^a	Meta-analysis		Analysis Conditional on the Coding Variant	
					Best Hit OR	Best Hit P	Best Hit OR	Best Hit P
A. Missense Coding Variants								
<i>PTPN22</i>	rs2476601	rs2476601	index	11	1,82 (1,73-1,92)	<1e-100	-	-
<i>CTLA4</i>	rs3087243	rs231775	0,51	11	0,88 (0,85-0,91)	5.09e-14	0,87 (0,82-0,91)	9.53e-09
<i>FCGR2A</i>	rs10494360	rs1801274	0,15	10	1,15 (1,09-1,21)	1.77e-07	1,11 (1,05-1,17)	0.0005
<i>TNFRSF14</i>	rs11585048	rs2234167 ^b	0,07	11	0,89 (0,85-0,92)	7.38e-11	0,89 (0,86-0,93)	3.57e-09
<i>TNFAIP3</i>	rs6920220	rs2230926	0,03	11	1,19 (1,15-1,24)	1.85e-18	1,17 (1,12-1,22)	2.74e-14
<i>CD2</i>	rs624988	rs699738 ^b	0,02	11	1,1 (1,07-1,14)	2,40E-08	1,09 (1,06-1,13)	3.24e-07
<i>TAGAP</i>	rs2451258	rs41267765	0,009	11	0,89 (0,86-0,93)	5.60e-10	0,9 (0,87-0,93)	2.95e-09
<i>TNFRSF14</i>	rs11585048	rs2234163 ^b	0,002	7	0,88 (0,85-0,92)	6.52e-09	0,88 (0,85-0,92)	1.61e-08
<i>BLK</i>	rs922483	rs55758736	0,001	11	1,09 (1,05-1,14)	1.81e-06	1,1 (1,06-1,14)	1.36e-06
B. Synonymous Coding Variants								
<i>FCGR2A</i>	rs10494360	rs11810143 ^b	0,99	10	1,15 (1,09-1,21)	1.77e-07	-	-
<i>PTPRC</i>	rs1932437	rs7540378	0,87	10	0,86 (0,82-0,91)	5.19e-08	0,86 (0,74-0,99)	0.043
<i>TRAF1</i>	rs10760126	rs2239657	0,7	11	1,12 (1,08-1,15)	1.97e-10	1,11 (1,05-1,18)	0.0007
<i>IL2RA</i>	rs10795791	rs2228150	0,032	11	1,11 (1,07-1,15)	2.91e-09	1,1 (1,06-1,14)	1.32e-07
<i>CCR6</i>	rs1571878	rs2071171	0,02	10	1,14 (1,10-1,19)	9.39e-14	1,14 (1,1-1,18)	9.67e-13
<i>CD58</i>	rs624988	rs35768283	0,003	10	1,1 (1,06-1,14)	5.76e-08	1,1 (1,06-1,14)	1.36e-07
<i>IL2RB</i>	rs3218251	rs228953	0,0011	11	1,1 (1,06-1,14)	1.11e-06	1,1 (1,06-1,14)	1.95e-06

^a Number of collections included in the analysis

^b Coding variants represented by a proxy in complete linkage disequilibrium

Table S6. Description of the Variants in Strong LD ($r^2 > 0.8$) with the Independent Signal of Association at CD2

SNP	Coordinates ^a	r^2 ^b	Annotation	Function / PolyPhen Prediction	GERP Conservation Score
rs798045	chr1:117300926	1	intron 2 of CD2	-	1.26
rs798044	chr1:117302613	1	intron 2 of CD2	-	-4.45
rs798040	chr1:117305283	0.973	intron 3 of CD2	-	-3.16
rs798039	chr1:117306131	0.973	intron 3 of CD2	-	0.149
rs798038	chr1:117307358	0.973	intron 4 of CD2	-	-3.22
rs798037	chr1:117308238	1	intron 4 of CD2	-	-6.53
rs798036	chr1:117308830	1	intron 4 of CD2	-	-3.79
rs798035	chr1:117310007	1	intron 4 of CD2	-	0.487
rs810048	chr1:117310296	1	intron 4 of CD2	-	-1.99
rs699738	chr1:117311147	index	exon 5 of CD2	missense / benign	-2.99
rs798034	chr1:117312119	1	269 bp 3' of CD2	-	-0.064
rs798033	chr1:117313078	1	1.23 kb 3' of CD2	-	-5.84
rs1088457	chr1:117316205	1	4.36 kb 3' of CD2	-	-0.451
rs798032	chr1:117317100	1	5.25 kb 3' of CD2	-	0.235
rs798031	chr1:117317129	1	5.28 kb 3' of CD2	-	0.235
rs813335	chr1:117318351	0.946	6.50 kb 3' of CD2	-	-0.503
rs1088456	chr1:117319616	0.946	7.77 kb 3' of CD2	-	-0.298

^a Coordinates in hg19

^b linkage disequilibrium (LD) with the missense variant rs699738